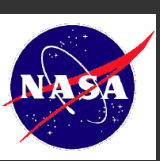


Data Access Services that Make Remote Sensing Data Easier to Use



Christopher Lynnes

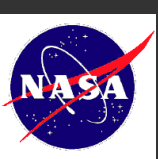
*Goddard Earth Sciences Data and Information
Center*



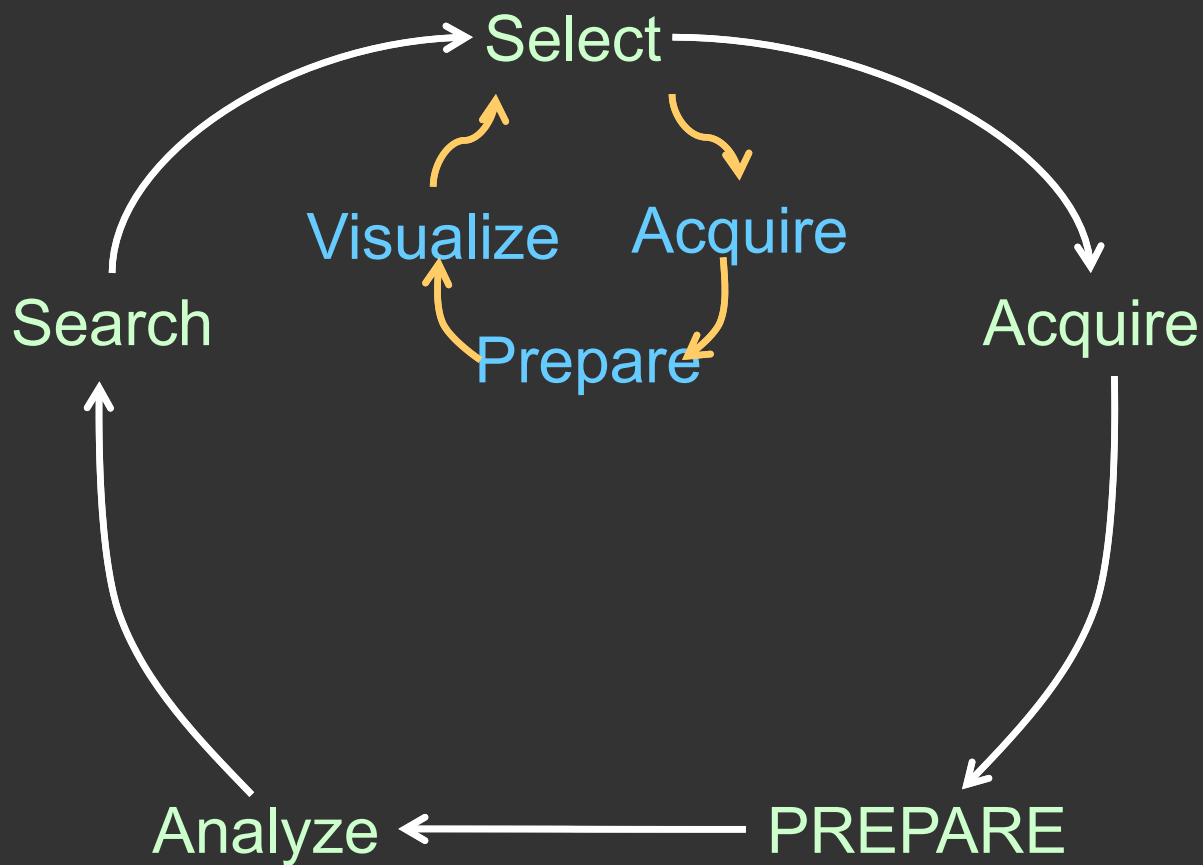
Goddard Earth Sciences Data and Information Services Center

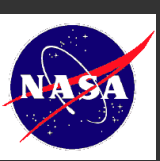


- ❧ GES DISC began as the Goddard Distributed Active Archive Center (DAAC)
 - ❧ Ingest, process, store and distribute Earth science data (mostly remote sensing)
- ❧ In the last decade, services have been added
 - ❧ Discovery
 - ❧ Access-related



The Data Usage Cycle



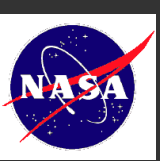


Preparation Steps

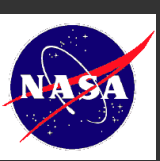


- ❧ Subsetting
 - ❧ Variable
 - ❧ Space
 - ❧ Time
- ❧ Gridding / (re)projection
- ❧ Reformatting to work in the analysis tools
- ❧ Quality Filtering

***How much of the Preparation process
can we build into the Access step?***



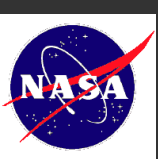
On-the-Fly Web Services: *executed on acquisition*



On-the-Fly Web Services



- ❧ REST-like: acquire as URLs
 - ❧ Limits error return possibilities
 - ❧ Requires an HTTP trick (shhh...) for long-running processes
- ❧ Accommodates any executable that...
 - ❧ ...Takes one file as input
 - ❧ ...Produces one file as output
- ❧ On-the-fly execution means minimal disk buffer requirements
 - ❧ No need to stage the whole request for pickup



On the Fly Subsetting



Subsetted Data Download Instructions:

- **Multiple** file download:
 - To use wget, you will need version 1.11.1 or later. To find the version, run 'wget -V'.
 - If necessary, [obtain the latest version of wget](#)
 - Download [the list of URLs](#)
 - Run: `wget --content-disposition -i wget_khElxSN`
- **Single** file download: click on the link for each file to be downloaded:
 - [MERRA100.prod.assim.inst3_3d_asm_Cp.19790131.SUB.nc](#)
 - [MERRA100.prod.assim.inst3_3d_asm_Cp.19790130.SUB.nc](#)
 - [MERRA100.prod.assim.inst3_3d_asm_Cp.19790129.SUB.nc](#)
 - [MERRA100.prod.assim.inst3_3d_asm_Cp.19790128.SUB.nc](#)
 - [MERRA100.prod.assim.inst3_3d_asm_Cp.19790127.SUB.nc](#)

Parameters

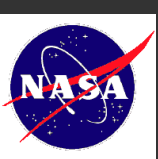
NOTE: Default Selection is All

- ☐ SLP = Sea-level pressure
- ☐ PS = Surface pressure
- ☐ PHIS = Surface Geopotential
- ☐ H = Geopotential Height
- ☐ O3 = Ozone mixing ratio
- ☐ QV = Specific humidity
- ☐ QL = Cloud liquid water mixing ratio
- ☐ QI = Cloud ice mixing ratio
- ☐ RH = Relative Humidity
- ☐ T = Air Temperature
- ☒ U = Eastward wind component
- ☒ V = Northward wind component
- ☐ EPV = Ertel potential vorticity
- ☐ OMEGA = Vertical pressure velocity

Additional Options

NOTE: Default Selection is All

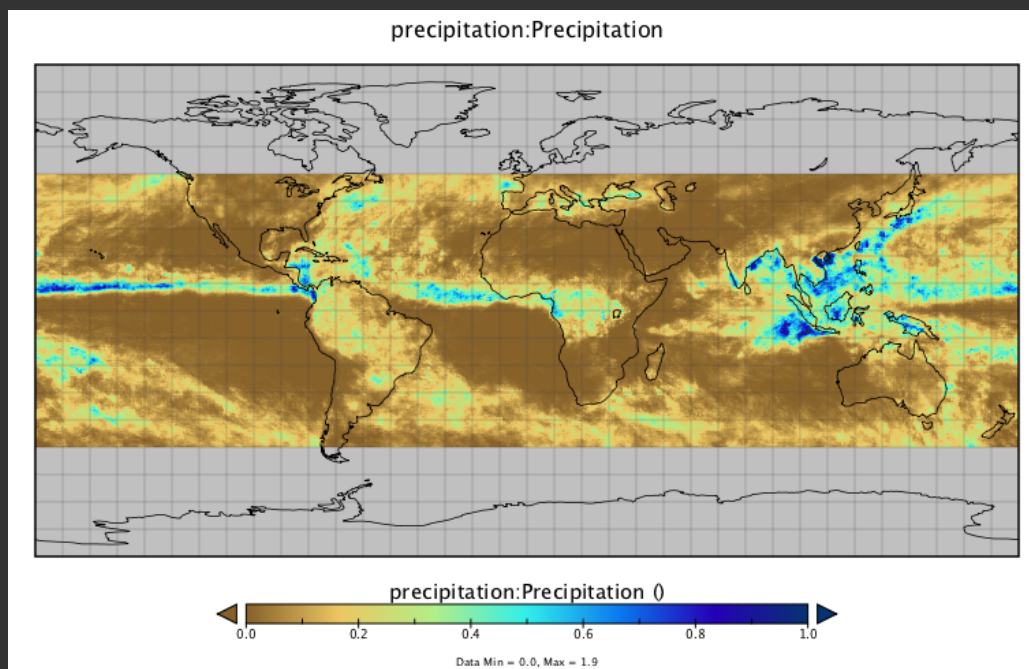
Time Subset



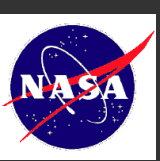
On-the-fly Conversion to netCDF (network Common Data Form)



- ☞ Most Earth Observing System datasets are in Hierarchical Data Format (HDF)
- ☞ BUT, many visualization tools understand netCDF “better”



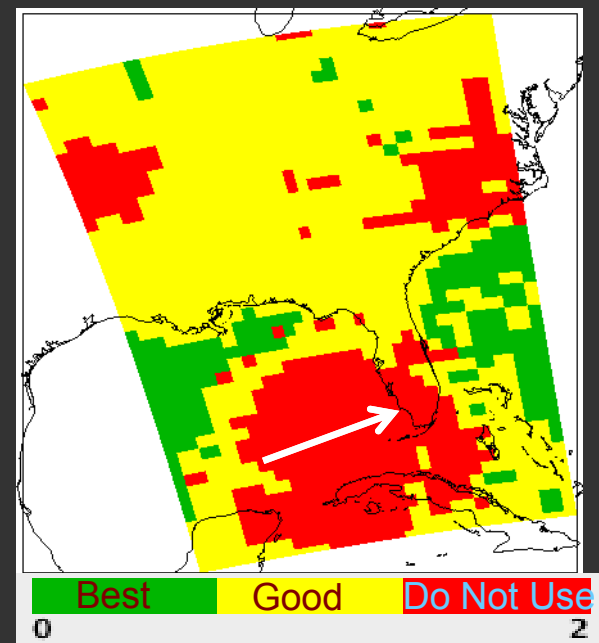
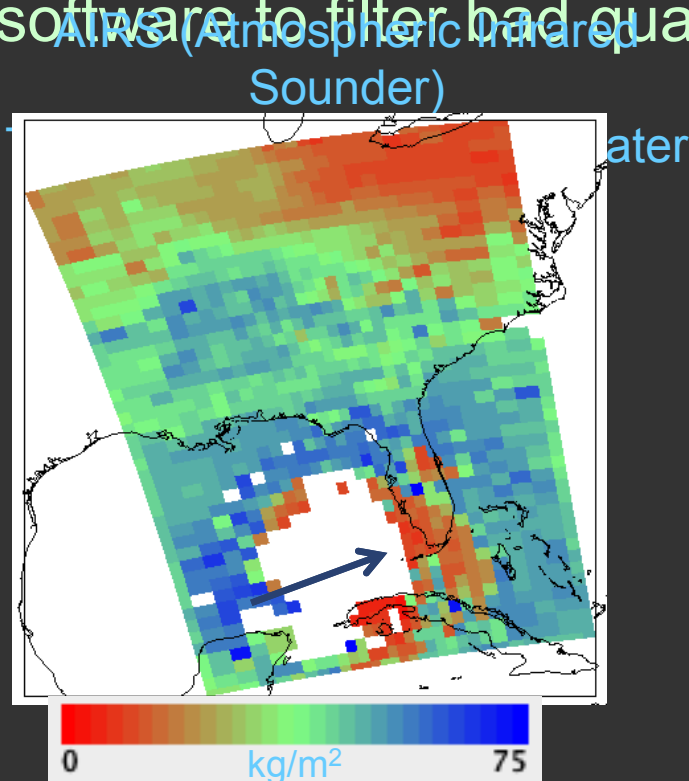
TRMM Monthly Rainfall rate for Oct 2011 in Panoply
<http://www.giss.nasa.gov/tools/panoply/>



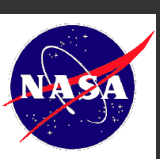
Data Quality Screening Service



- Level 2 Satellite data often comes with quality control flags
- Until now, each user typically had to write his/her own software to filter bad quality data—or ignore them



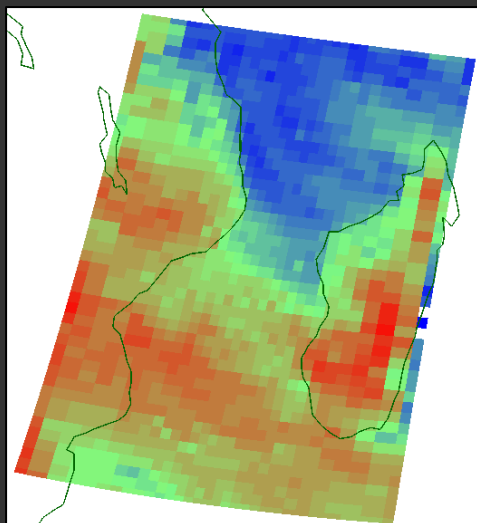
Hurricane Ike, 9/10/2008



The Data Quality Screening Service for AIRS Level 2 swath data



Original data
array

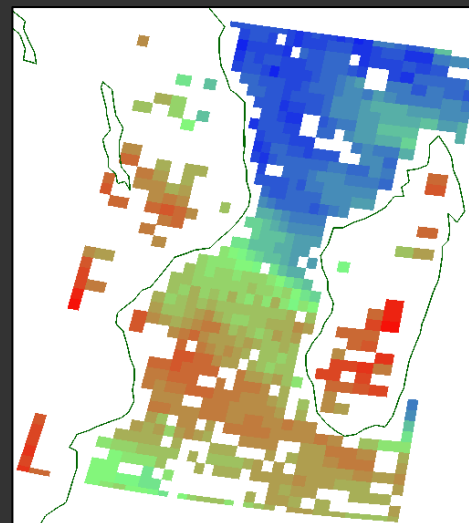


Total column
precipitable water

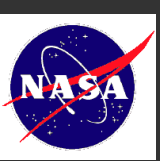
Mask based on
user criteria
 $\text{Quality flag} < 2$



Good quality
data pixels
retained

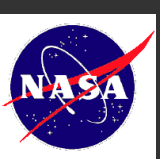


Output file has the same format and structure as the
input file, with fill values replacing the low-quality data

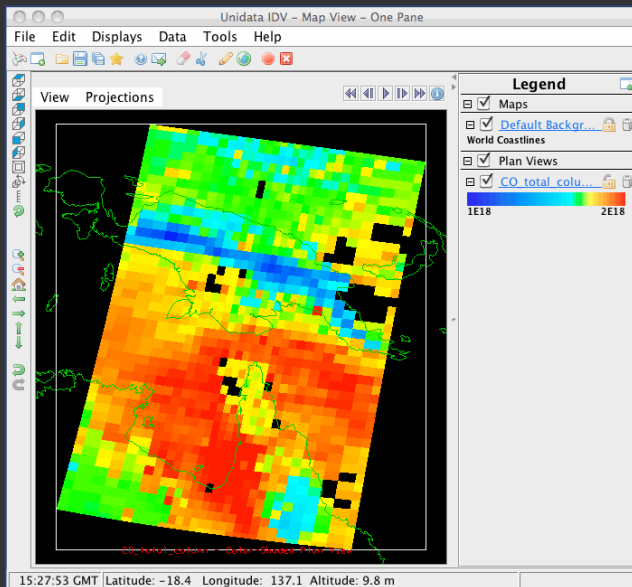


OPeNDAP*: *a protocol standard for remote access*

*Open-source Project for a Network Data Access Protocol



OPeNDAP: Subsetting and more



← Data Attributes

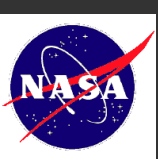
← Data Objects

Server

Data

- ☞ Subsetting
 - ☞ individual variables
 - ☞ slices of variables

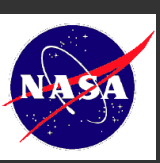
- ☞ Reformatting: download as...
 - ☞ ASCII
 - ☞ netCDF



Varieties of OPeNDAP

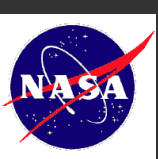


- ❧ Hyrax
 - ❧ High performance
 - ❧ Reformat to netCDF
- ❧ GrADS Data Server
 - ❧ Multiple input formats
 - ❧ Server-side processing
- ❧ THREDDS Data Server
 - ❧ Aggregation
 - ❧ Web Coverage Service, netCDF Subsetter
- ❧ Others: ERDDAP, PyDAP, Dapper...



Giovanni:

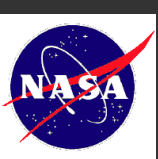
online analysis and visualization



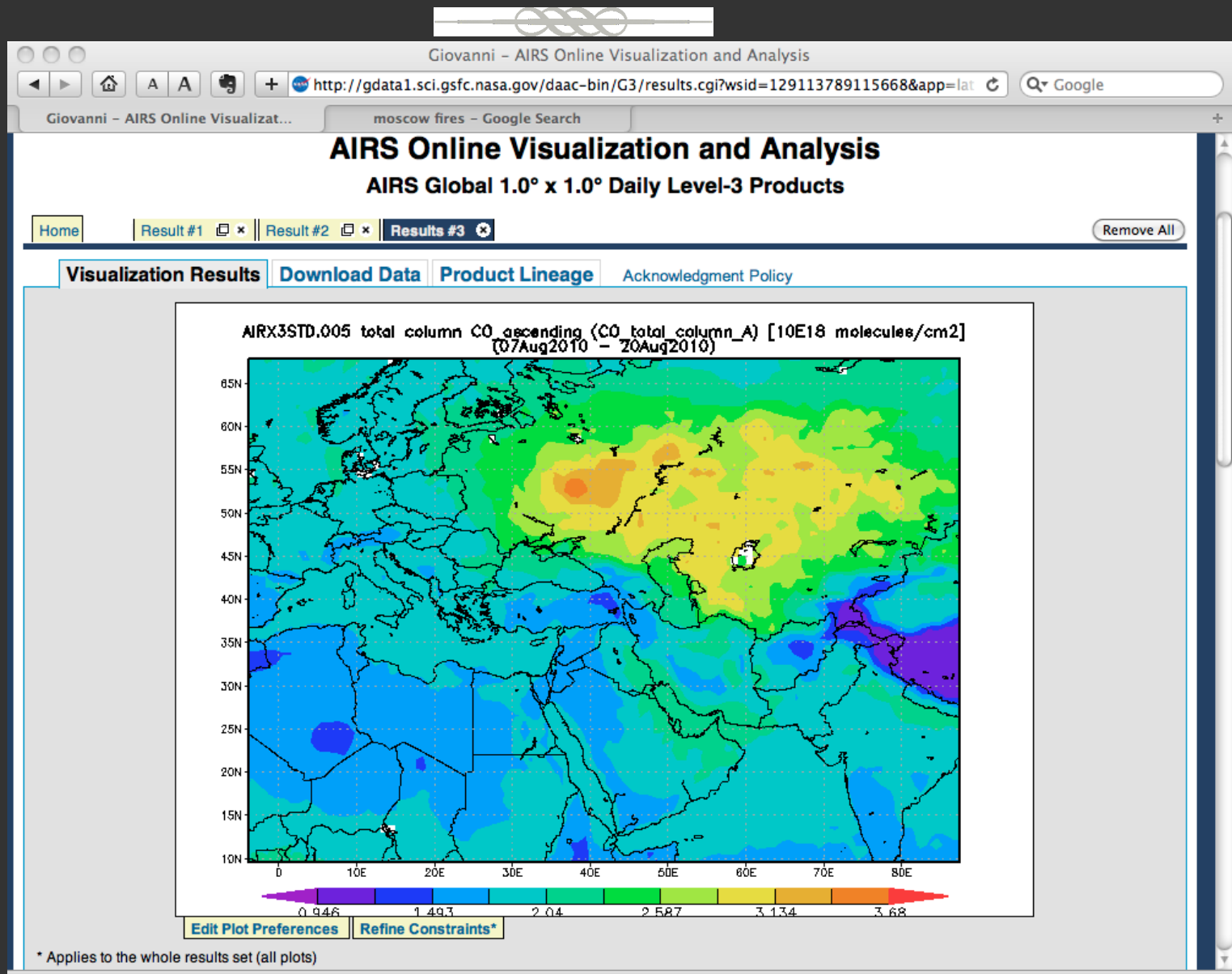
Giovanni

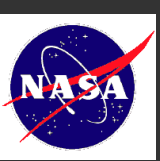


- ❧ Analysis and visualization server
- ❧ Workflow paradigm
- ❧ Steps for:
 - ❧ Fetching
 - ❧ Subsetting
 - ❧ Quality filtering
 - ❧ Regridding
 - ❧ Averaging
 - ❧ Visualization
- ❧ Output can be downloaded

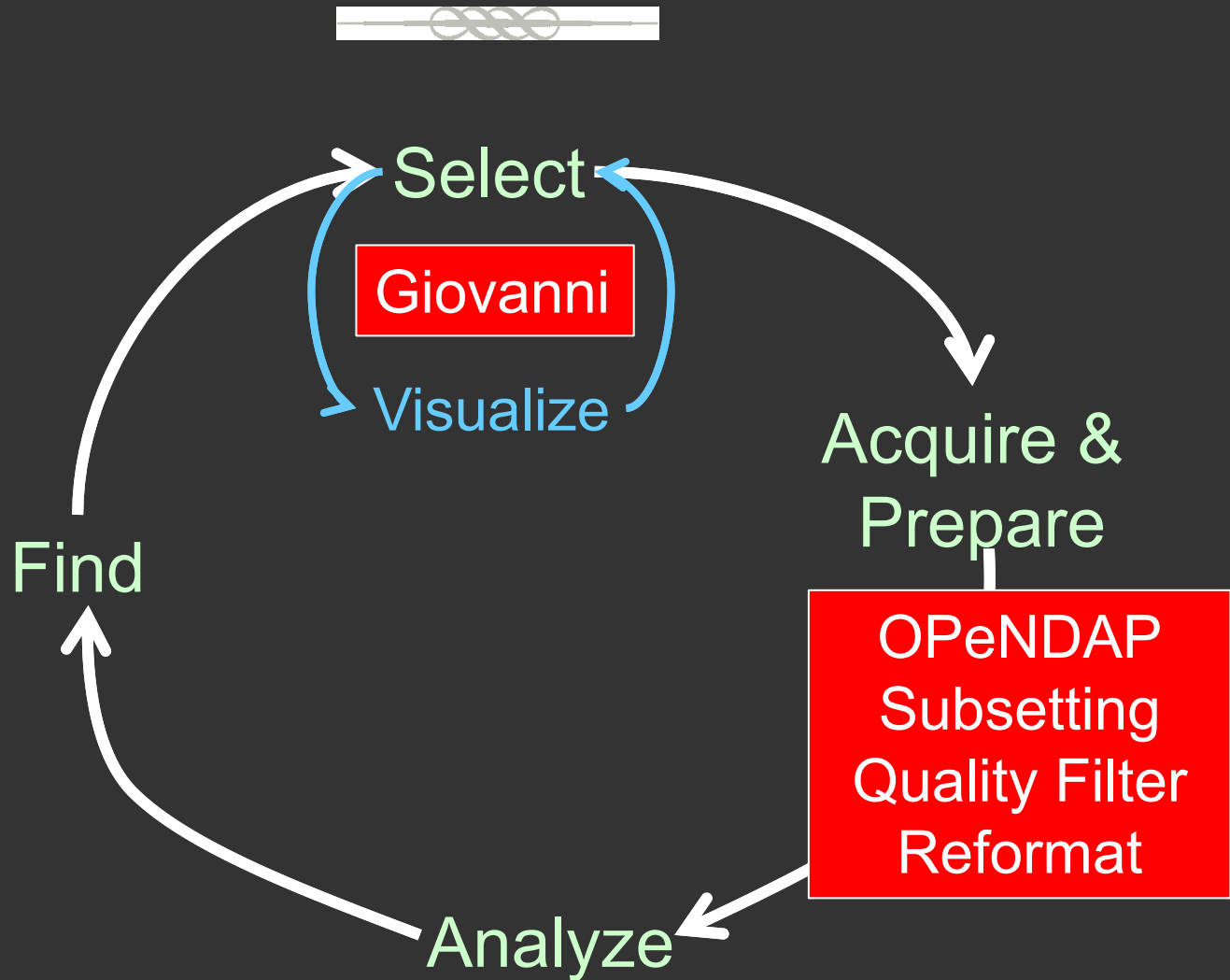


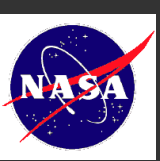
Example: Carbon Monoxide from 2010 Russian wildfires



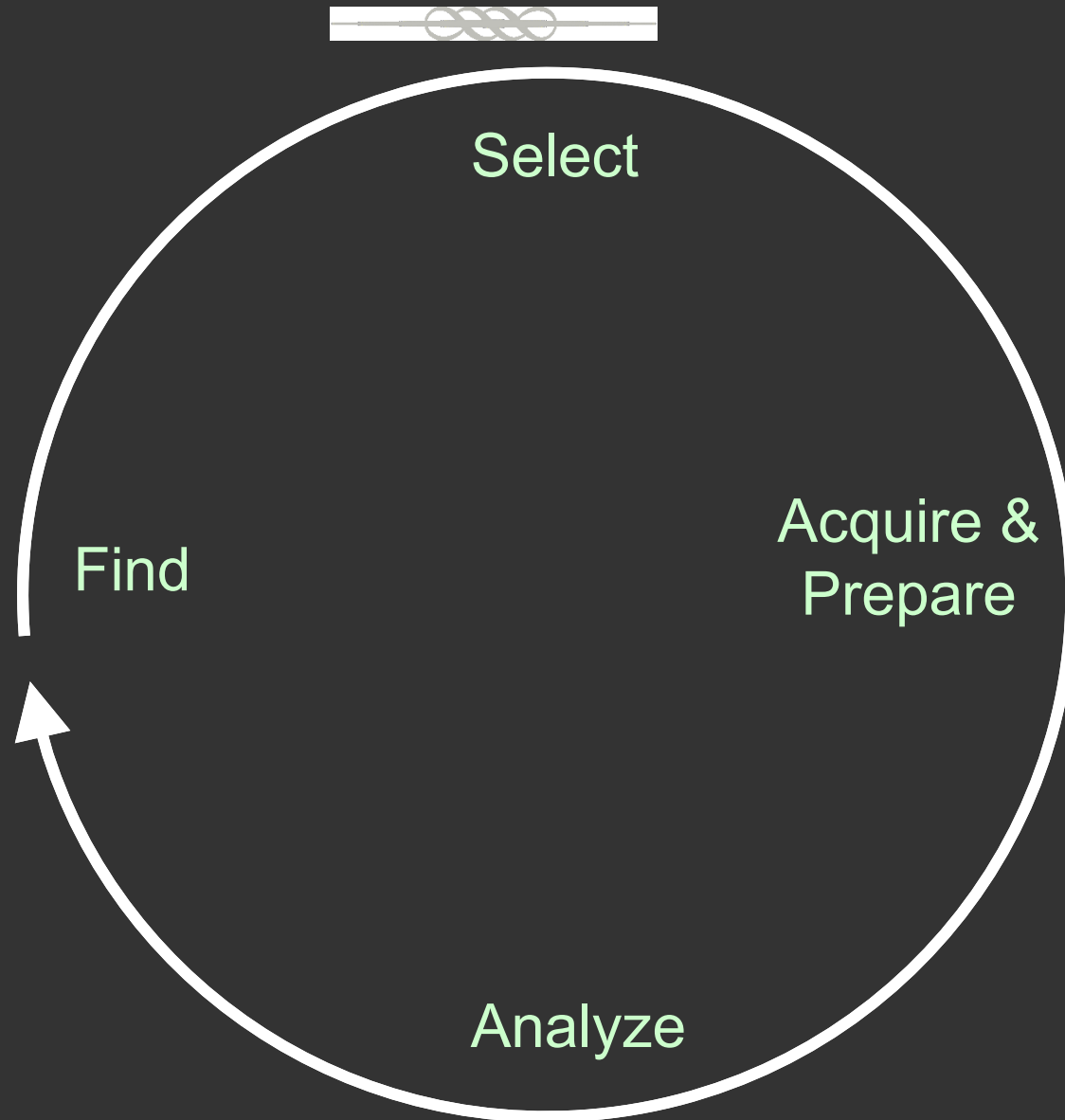


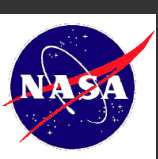
The Data Usage Cycle Refactored





Frontier: Seamless interaction of steps

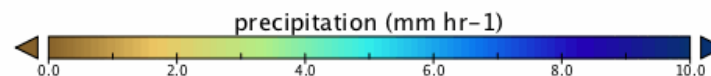
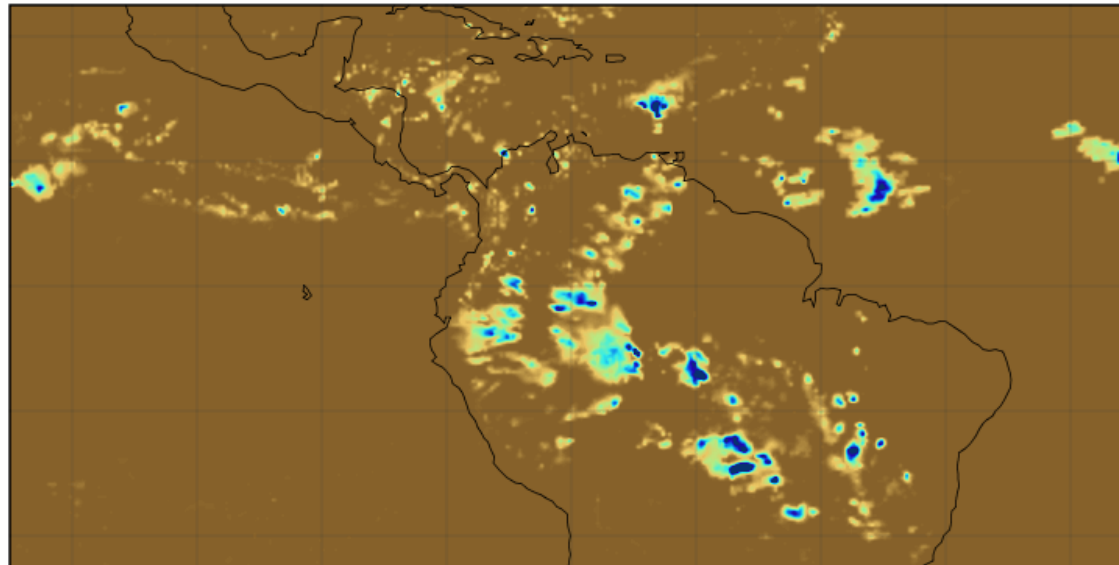




Seamless Search and Analysis



precipitation



Start: 2010-10-31 00:00:00

End: 2010-10-31 23:59:59

Area:

-115

22.5

-25

-22.5

Measurement: Soil Moisture (SMAP)

Filter Quality?:

X

Fetch